

Claims

WHAT IS CLAIMED IS:

1. A method of determining a field-weighted score for a document having multiple fields relative to a query having a plurality of query terms, the method comprising:

replicating each field of the document in accordance with a field weight corresponding to the field to produce an individual field set corresponding to each field in the document;

combining each field set for the document into a virtual document;

indexing the virtual document to produce a virtual document statistics; and

computing the field-weighted score from the virtual document index based on the query.

2. The method of claim 1 wherein the query is associated with a search and the field-weighted score represents a level of relevance of the document to the query.

3. The method of claim 1 wherein each field weight is represented by an integer value and the replicating operation comprises:

generating each field set to include a number of copies of a field of the document, wherein the number of copies equals the integer value.

4. The method of claim 1 wherein the replicating operation comprises:

concatenating copies of one of the fields into a field set.

1 5. The method of claim 1 wherein the combining operation comprises:
2 concatenating each field set into the virtual document.

3 6. The method of claim 1 wherein the computing operation comprises:
4 computing a field-weighted document weight for each query term in the
5 query from the virtual document statistics.

6 7. The method of claim 1 wherein the computing operation comprises:
7 computing a field-weighted document weight for each query term in the
8 query from the virtual document statistics; and
9 computing the field-weighted score based on the field-weighted document
10 weight for each query term.

11 8. The method of claim 1 further comprising:
12 ranking the field-weighted score with field-weighted scores of other
13 documents.
14

1 9. A computer program product encoding a computer program for
2 executing on a computer system a computer process for determining a field-
3 weighted score for a document having multiple fields relative to a query having a
4 plurality of query terms, the computer process comprising:

5 replicating each field of the document in accordance with a field weight
6 corresponding to the field to produce an individual field set corresponding to each
7 field in the document;

8 combining each field set for the document into a virtual document;

9 indexing the virtual document to produce a virtual document statistics; and

10 computing the field-weighted score from the virtual document index based
11 on the query.

12
13 10. The computer program product of claim 9 wherein the query is
14 associated with a search and the field-weighted score represents a level of
15 relevance of the document to the query.

16 11. The computer program product of claim 9 wherein each field weight is
17 represented by an integer value and the replicating operation comprises:

18 generating each field set to include a number of copies of a field of the
19 document, wherein the number of copies equals the integer value.

20
21 12. The computer program product of claim 9 wherein the replicating
22 operation comprises:

23 concatenating copies of one of the fields into a field set.

1 13. The computer program product of claim 9 wherein the combining
2 operation comprises:

3 concatenating each field set into the virtual document.

4 14. The computer program product of claim 9 wherein the computing
5 operation comprises:

6 computing a field-weighted document weight for each query term in the
7 query from the virtual document statistics.

8 15. The computer program product of claim 9 wherein the computing
9 operation comprises:

10 computing a field-weighted document weight for each query term in the
11 query from the virtual document statistics; and

12 computing the field-weighted score based on the field-weighted document
13 weight for each query term.

14 16. The computer program product of claim 9 further comprising:

15 ranking the field-weighted score with field-weighted scores of other
16 documents.
17
18
19
20
21
22
23
24
25

1 17. A method of determining a field-weighted score for a document having
2 multiple fields relative to a query having a plurality of query terms, the method
3 comprising:

4 determining a field-specific term frequency for each field in the document
5 for each query term;

6 weighting each field-specific term frequency according to a field weight
7 designated for the corresponding field to compute a field-weighted term frequency
8 for each query term;

9 computing a field-weighted document weight for each query term based on
10 the field-weighted term frequency for each query term; and

11 computing the field-weighted score as a function of the field-weighted
12 document weight of all query terms.

13
14 18. The method of claim 17 wherein the query is associated with a search
15 and the field-weighted score represents a level of relevance of the document to the
16 query.

17 19. The method of claim 17 further comprising:

18 computing a field-weighted document length based on a field weight for
19 each field and a field length for each field, wherein the operation of computing a
20 field-weighted document weight comprises computing a field-weighted document
21 weight for each query term based on the field-weight term frequency for each
22 query term and the field-weighted document length.

1 20. The method of claim 17 further comprising:

2 computing a field-weighted document length based on a field weight for
3 each field and a field length for each field by summing one or more weighted field
4 lengths of the fields in the document, each weighted field length being a field
5 length weighted by a corresponding field weight.

6 21. The method of claim 17 further comprising:

7 computing a field-weighted document length based on a field weight for
8 each field and a field length for each field by summing one or more weighted field
9 lengths of the fields in the document, each weighted field length being a field
10 length weighted by a corresponding field weight, wherein the operation of
11 computing a field-weighted document weight comprises computing a field-
12 weighted document weight for each query term based on the field-weight term
13 frequency for each query term and the field-weighted document length.

14 22. The method of claim 17 wherein the determining operation comprises:

15 determining the field-specific term frequency for each field from document
16 statistics associated with the document, the document statistics including a field-
17 weighted term frequency for at least one query term in the document.

18 23. The method of claim 17 wherein the determining operation comprises:

19 determining the field length for each field from document statistics
20 associated with the document.
21
22
23
24
25

1 24. The method of claim 17 wherein the operation of computing a field-
2 weighted document weight comprises:

3 summing one or more weighted field-specific term frequencies of the fields
4 in the document.

5 25. The method of claim 17 further comprising:

6 ranking the field-weighted score with field-weighted scores of other
7 documents.

1 26. A computer program product encoding a computer program for
2 executing on a computer system a computer process for determining a field-
3 weighted score for a document having multiple fields relative to a query having a
4 plurality of terms, the computer process comprising:

5 determining a field-specific term frequency for each field in the document
6 for each query term;

7 weighting each field-specific term frequency according to a field weight
8 designated for the corresponding field to compute a field-weighted term frequency
9 for each query term;

10 computing a field-weighted document weight for each query term based on
11 the field-weighted term frequency for each query term; and

12 computing the field-weighted score as a function of the field-weighted
13 document weight of all query terms.

14
15 27. The computer program product of claim 26 wherein the query is
16 associated with a search and the field-weighted score represents a level of
17 relevance of the document to the query.

18 28. The computer program product of claim 26 wherein the computer
19 process further comprises:

20 computing a field-weighted document length based on a field weight for
21 each field and a field length for each field, wherein the operation of computing a
22 field-weighted document weight comprises computing a field-weighted document
23 weight for each query term based on the field-weight term frequency for each
24 query term and the field-weighted document length.
25

1 29. The computer program product of claim 26 wherein the computer
2 process further comprises:

3 computing a field-weighted document length based on a field weight for
4 each field and a field length for each field by summing one or more weighted field
5 lengths of the fields in the document, each weighted field length being a field
6 length weighted by a corresponding field weight.

7 30. The computer program product of claim 26 wherein the computer
8 process further comprises:

9 computing a field-weighted document length based on a field weight for
10 each field and a field length for each field by summing one or more weighted field
11 lengths of the fields in the document, each weighted field length being a field
12 length weighted by a corresponding field weight, wherein the operation of
13 computing a field-weighted document weight comprises computing a field-
14 weighted document weight for each query term based on the field-weight term
15 frequency for each query term and the field-weighted document length.

16
17 31. The computer program product of claim 26 wherein the determining
18 operation comprises:

19 determining the field-specific term frequency for each field from document
20 statistics associated with the document, the document statistics including a field-
21 weighted term frequency for at least one query term in the document.

1 32. The computer program product of claim 26 wherein the determining
2 operation comprises:

3 determining the field length for each field from document statistics
4 associated with the document.

5 33. The computer program product of claim 26 wherein the operation of
6 computing a field-weighted document weight comprises:

7 summing one or more weighted field-specific term frequencies of the fields
8 in the document.

9 34. The computer program product of claim 26 further comprising:
10

11 ranking the field-weighted score with field-weighted scores of other
12 documents.

1 35. A system for determining a field-weighted score for a document having
2 multiple fields relative to a query having a plurality of terms, the system
3 comprising:

4 a field-weighted term frequency calculator that determines a field-specific
5 term frequency for each field in the document for each query term and weights
6 each field-specific term frequency according to a field weight identified for the
7 corresponding field to compute a field-weighted term frequency for each query
8 term;

9 a field-weighted document weight calculator that computes a field-
10 weighted document weight for each query term based on the field-specific term
11 frequency for each query term; and

12 a document score calculator that computes the field-weighted score as a
13 function of the field-weighted document weight of all query terms.

14
15 36. The system of claim 35 wherein the query is associated with a search
16 and the field-weighted score represents a level of relevance of the document to the
17 query.

18 37. The system of claim 35 further comprising:

19 a field-weighted document length calculator that computes a field-weighted
20 document length based on a field weight for each field and a field length for each
21 field, wherein the field-weighted document weight calculator computes a field-
22 weighted document weight for each query term based on the field-weight term
23 frequency for each query term and the field-weighted document length.
24
25

1 38. A method of determining a field-weighted score for a document having
2 multiple fields relative to a query having a plurality of query terms, the method
3 comprising:

4 computing a field-weighted term frequency for each query term based on
5 field weights designated for individual fields in the document;

6 computing a field-weighted document weight for each query term based on
7 the field-weighted term frequency for each field in the document; and

8 computing the field-weight score as a function of the field-weighted
9 document weights of the query terms.
10

11 39. The method of claim 38 further comprising:

12 computing a field-weighted document length based on a field weight for
13 each field and a field length for each field, wherein the operation of computing a
14 field-weighted document weight comprises computing a field-weighted document
15 weight for each query term based on the field-weight term frequency for each
16 query term and the field-weighted document length.

17 40. The method of claim 38 wherein computing a field-weighted document
18 weight comprises:

19 computing the field-weighted document weight using a field-weighted free
20 parameter of a BM25 document weighting function, the field-weighted free
21 parameter being based on a corresponding optimized free parameter computed in a
22 non-field-weighted configuration.
23
24
25

1 41. The method of claim 38 wherein computing a field-weighted document
2 weight comprises:

3 computing the field-weighted document weight using a field-weighted free
4 parameter of a BM25 document weighting function, the field-weighted free
5 parameter being based on an average term frequency over all terms in a non-field-
6 weighted configuration, an average term frequency over all terms in a field-
7 weighted configuration, and a corresponding optimized free parameter computed
8 in the non-field-weighted configuration.

9 42. The method of claim 38 wherein computing a field-weighted document
10 weight comprises:

11 computing the field-weighted document weight using a factor reflecting a
12 dependence on a number of the fields in the document in which a query term
13 occurs.

14 43. The method of claim 38 wherein computing a field-weighted score
15 comprises:

16 computing the field-weighted score using a factor reflecting a dependence
17 on which field in the document includes the most query terms.
18
19
20
21
22
23
24
25

1 44. A computer program product encoding a computer program for
2 executing on a computer system a computer process for determining a field-
3 weighted score for a document having multiple fields relative to a query having a
4 plurality of query terms, the computer process comprising:

5 computing a field-weighted term frequency for each query term based on
6 field weights designated for individual fields in the document;

7 computing a field-weighted document weight for each query term based on
8 the field-weighted term frequency for each field in the document; and

9 computing the field-weight score as a function of the field-weighted
10 document weights of the query terms.

11
12 45. The computer program product of claim 44 wherein the computer
13 process further comprises:

14 computing a field-weighted document length based on a field weight for
15 each field and a field length for each field, wherein the operation of computing a
16 field-weighted document weight comprises computing a field-weighted document
17 weight for each query term based on the field-weight term frequency for each
18 query term and the field-weighted document length.

19 46. The computer program product of claim 44 wherein computing a field-
20 weighted document weight comprises:

21 computing the field-weighted document weight using a field-weighted free
22 parameter of a BM25 document weighting function, the field-weighted free
23 parameter being based on a corresponding optimized free parameter computed in a
24 non-field-weighted configuration.
25

1 47. The computer program product of claim 44 wherein computing a field-
2 weighted document weight comprises:

3 computing the field-weighted document weight using a field-weighted free
4 parameter of a BM25 document weighting function, the field-weighted free
5 parameter being based on an average term frequency over all terms in a non-field-
6 weighted configuration, an average term frequency over all terms in a field-
7 weighted configuration, and a corresponding optimized free parameter computed
8 in the non-field-weighted configuration.

9 48. The computer program product of claim 44 wherein computing a field-
10 weighted document weight comprises:

11 computing the field-weighted document weight using a factor reflecting a
12 dependence on a number of the fields in the document in which a query term
13 occurs.

14 49. The computer program product of claim 44 wherein computing a field-
15 weighted score comprises:

16 computing the field-weighted score using a factor reflecting a dependence
17 on which field in the document includes the most query terms.
18
19
20
21
22
23
24
25

1 50. A system for determining a field-weighted score for a document having
2 multiple fields relative to a query having a plurality of query terms, the system
3 comprising:

4 a field-weighted term frequency calculator that computes a field-weighted
5 term frequency for each query term based on field weights designated for
6 individual fields in the document;

7 a field-weighted document weight calculator that computes a field-
8 weighted document weight for each query term based on the field-weighted term
9 frequency for each field in the document; and

10 a search engine that computes the field-weighted score as a function of the
11 field-weighted document weights of the query terms.

12
13 51. The system of claim 50 further comprising:

14 a field-weighted document length calculator that computes a field-weighted
15 document length based on a field weight for each field and a field length for each
16 field, wherein the field-weighted document weight calculator computes a field-
17 weighted document weight for each query term based on the field-weight term
18 frequency for each query term and the field-weighted document length.